

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/511,436A
Source: PU710
Date Processed by STIC: 1/5/06

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:15

Input Set : A:\81356224.APP

Output Set : N:\CRF4\01052006\J511436A.raw

3 <110> APPLICANT: KOBAYASHI, KAZUO
 4 KITAGAWA, YOSHINORI
 5 KOMEDA, TOSHIHIRO
 6 KAWASHIMA, NAGAKO
 7 JIGAMI, YOSHIFUMI
 8 CHIBA, YASUNORI
 10 <120> TITLE OF INVENTION: METHYLOTROPH PRODUCING MAMMALIAN TYPE SUGAR CHAIN
 12 <130> FILE REFERENCE: 081356-0224
 14 <140> CURRENT APPLICATION NUMBER: 10/511,436A
 15 <141> CURRENT FILING DATE: 2004-10-25
 17 <150> PRIOR APPLICATION NUMBER: PCT/JP03/05464
 18 <151> PRIOR FILING DATE: 2003-04-28
 20 <150> PRIOR APPLICATION NUMBER: JP 2002-127677
 21 <151> PRIOR FILING DATE: 2002-04-26
 23 <160> NUMBER OF SEQ ID NOS: 121
 25 <170> SOFTWARE: PatentIn Ver. 3.3
 27 <210> SEQ ID NO: 1
 28 <211> LENGTH: 11
 29 <212> TYPE: PRT
 30 <213> ORGANISM: Saccharomyces cerevisiae
 32 <400> SEQUENCE: 1
 33 Ala Tyr Met Phe Lys Tyr Asp Ser Thr His Gly
 34 1 5 10
 37 <210> SEQ ID NO: 2
 38 <211> LENGTH: 11
 39 <212> TYPE: PRT
 40 <213> ORGANISM: Saccharomyces cerevisiae
 42 <400> SEQUENCE: 2
 43 Asp Gly Pro Ser His Lys Asp Trp Arg Gly Gly
 44 1 5 10
 47 <210> SEQ ID NO: 3
 48 <211> LENGTH: 32
 49 <212> TYPE: DNA
 50 <213> ORGANISM: Artificial Sequence
 52 <220> FEATURE:
 53 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 54 primer PGP5 for amplification of 5'-region of
 55 Ogataea minuta GAP gene
 57 <220> FEATURE:
 58 <221> NAME/KEY: modified_base
 59 <222> LOCATION: (3)
 60 <223> OTHER INFORMATION: a, c, g or t
 62 <220> FEATURE:

RAW SEQUENCE LISTING

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:15

Input Set : A:\81356224.APP

Output Set: N:\CRF4\01052006\J511436A.raw

```

63 <221> NAME/KEY: modified_base
64 <222> LOCATION: (24)
65 <223> OTHER INFORMATION: a, c, g or t
67 <220> FEATURE:
68 <221> NAME/KEY: modified_base
69 <222> LOCATION: (27)
70 <223> OTHER INFORMATION: a, c, g or t
72 <400> SEQUENCE: 3
W--> 73 gcntayatgt tyaartayga ywsnacncay gg 32
76 <210> SEQ ID NO: 4
77 <211> LENGTH: 32
78 <212> TYPE: DNA
79 <213> ORGANISM: Artificial Sequence
81 <220> FEATURE:
82 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
83     primer PGP3 for amplification of 3'-region of
84     Ogataea minuta GAP gene
86 <220> FEATURE:
87 <221> NAME/KEY: modified_base
88 <222> LOCATION: (3)
89 <223> OTHER INFORMATION: a, c, g or t
91 <220> FEATURE:
92 <221> NAME/KEY: modified_base
93 <222> LOCATION: (6)
94 <223> OTHER INFORMATION: a, c, g or t
96 <220> FEATURE:
97 <221> NAME/KEY: modified_base
98 <222> LOCATION: (21)
99 <223> OTHER INFORMATION: a, c, g or t
101 <220> FEATURE:
102 <221> NAME/KEY: modified_base
103 <222> LOCATION: (24)
104 <223> OTHER INFORMATION: a, c, g or t
106 <220> FEATURE:
107 <221> NAME/KEY: modified_base
108 <222> LOCATION: (27)
109 <223> OTHER INFORMATION: a, c, g or t
111 <400> SEQUENCE: 4
W--> 112 ccnccnckcc artcyttrtg nswnggnccr tc 32
115 <210> SEQ ID NO: 5
116 <211> LENGTH: 3186
117 <212> TYPE: DNA
118 <213> ORGANISM: Ogataea minuta
120 <400> SEQUENCE: 5
121 aagctttact ggttcaaggg gttaagtagg ggcgcggtct ggtctttgtg gttgtttcta 60
122 caccgaccac agttgacagc atcgactgct catcgaaaac ggtcgcagtg cggcaatctg 120
123 ctctatctaa tcccaggcta ctcgatccct gcacaacctc cagagtgatc cgaccgcact 180
124 gcccgagatt cagcagatc tcgcagcgca gcgtgcgttt taatccctca aatcaaggct 240
125 gtgcagaccc ggaggatgtg aagctgggac ggcgggaggg aagtctggag tggtagagaga 300

```

RAW SEQUENCE LISTING

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:15

Input Set : A:\81356224.APP

Output Set: N:\CRF4\01052006\J511436A.raw

```

126 atgtgggagc tgtgcaaagg ggcaatggtc actcagcgca gagcgatggt ggcgcggggg 360
127 ccaatatctc ggcaacaaga acgcccagg acgacgggac tctgaatgcg agcacgttgt 420
128 ctttcagaca gtccaccggt attccaatat tcgcaggact cgcgctcaga aacgcaaccc 480
129 cggcagattc gcgtccagtc aggccatctg cggcgagctg ctgcgctcgc gggctgcgcc 540
130 acaacgcata gccacatata cgtcaccgcc cgcccgtgg caacctgagg tttttccgca 600
131 acgggtgcac tgattgctgc gttaacgagg caactggaga tgtcagaggc caagtggagc 660
132 catatcacag cggactgcgc atctctggcc tgccggacgc ggtagcgtcc cgtctttttg 720
133 cggacagctt cttaaaacct ggctgaaact aagcgagacc tgcgacctgg aacgcccgc 780
134 caccctgaca cctccggagt tgtatcctca gaagcggagt aacctgcagg cctacgcaag 840
135 aaaagagccc gggacccatc gaccggaaaa gagggggtga gctagtgggg tagccttgg 900
136 gcagacctgg ggcagacctg ggtagtagacc agggccgaaa agggtcagag gaatcagggt 960
137 ggcacggcag tctataccgt agaagctctt ctgcacagca gcgagcagaa actgcacaga 1020
138 ggtccgttcg ccagtctcgt accaccaccg catgacccaa tcagcattga tgctccaca 1080
139 tgggtagtgc gcgcgaacgc ctggcaccca aacacaccac ttacgcttcc cgcaccgcgg 1140
140 tggttaacac tggcccggag tagtcatata cggagatttt ggcatgattc taattccggg 1200
141 tcgggacacg acctaaagtg cgtgcaaagc tcgggggcta aatgtttccc ggcgctcgcg 1260
142 gcgactcttg tgcgcgcccg cggcggttcg cgggagacgg gggaaagaga ggggtgaccg 1320
143 cagcgagcga tgggtgtgcca gatctcaggc cgagtcaaga caatatataa agagaggatt 1380
144 gtccactttt ctccaatagt atttgaccgg ggttgctctc tgttgatttt ttctagatca 1440
145 tacaattatt gtttgaattc actcaattaa catacacaaa tacaatacaa aatggcttac 1500
146 aacgtcggta tcaacggatt cggaagaatt ggtagactcg ttcttagaat tgctttgtcc 1560
147 agaaaggaca tcaacgtggt tgccgtgaat gatccattca tcgctgccga gtacgctgct 1620
148 tacatgttca agtacgactc cactcacgga agataccaag gtgaagtcac cttcgaggga 1680
149 aagtaccttg tgatcgacgg tcagaagatt gaggtgttcc aagagagaga ccctgctgac 1740
150 atcccatggg gtaaggaggg cgttgacttt gtcattgact ccaccggtgt gttcaccacc 1800
151 accgcccggc ctcaaaagca cattgatgct ggtgccaaga aggttatcat cactgctcca 1860
152 tccgctgacg ctccaatgtt cggtatgggt gtcaaccaca aggagtacac caaggacttg 1920
153 tccattgtct ccaacgttcc ctgtaccacc aactgtctgg ctccattggc caaggttgtt 1980
154 aacgacgttt tcggtattga gtctggtttg atgaccaccg tccactctat cactgccacc 2040
155 caaaagaccg ttgacggtcc atcccacaag gactggagag gaggaagaac cgcttcgggt 2100
156 aacatcattc catctccac cggtgccgtt aaggctgtcg gtaaggctct gccagctctt 2160
157 gctggtaagt tgactggtat gtctctgaga gttcctacca ccgatgttcc cgttggtgac 2220
158 ttgactgtca acttgaagac cccaaccacc tacgcagaga tctccgcgcg catcaagaag 2280
159 gcctctgagg gtgaacttgc cggatatctg gggtacactg aggacgcgct tgtctccact 2340
160 gacttcttga ccgacaacag atcttcgata tttagacgct ctgccggtat cttggtgacc 2400
161 ccaactttcg tcaagttgat ctcttggtac gataacgagt acggttactc caccagagtt 2460
162 gtcgacttgc ttgagcacgt tgccaaggte tcttcgctt aagtggatag atgaccaatg 2520
163 gcctctttta gtaaacattt cgttttgaat atatttcaag ttgaataatg aaagccttgt 2580
164 tgtagactta ctccgaagct ccggggcttc ggctccctga atttattttt tacatctctg 2640
165 caccgaaaaa ctggctattt gaaaaatttc gacgttttgc ttgaaactcg agttgaggag 2700
166 cattgcaaaa ttcgatcggt ttctaaccgga cgccagtcga gttattgtta tgtcacgtga 2760
167 catcaattgt cctctattcc tttttggccg atctcgtttg tgcgacggc ctccgaacag 2820
168 ttacttctac cggcagggat tggggatgat cgggatcgat gtcctcaact ccagaggctg 2880
169 atccgatgcg gtgggacttc atgcgtccaa atctgttggg tgatgtgctc ttctgctttt 2940
170 ttggtgacca aacgagatga caattgactg cattgaaaag gttattagct tttttggtct 3000
171 tctctgtgtg cgattcgagc ggtaccgtag gtaggtctgc tatggaggca tgcgtcataa 3060
172 gtcagccttg attaaccttc ggagctgcgc gatccacatc tctgcaccgc gcggaggcct 3120
173 ttgactgcag cattttaatt aatctcgtaa aataagctct taaacgagat tagcttacgg 3180
174 ggatcc

```

RAW SEQUENCE LISTING

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:15

Input Set : A:\81356224.APP

Output Set: N:\CRF4\01052006\J511436A.raw

```

177 <210> SEQ ID NO: 6
178 <211> LENGTH: 336
179 <212> TYPE: PRT
180 <213> ORGANISM: Ogataea minuta
182 <400> SEQUENCE: 6
183 Met Ala Tyr Asn Val Gly Ile Asn Gly Phe Gly Arg Ile Gly Arg Leu
184   1           5           10          15
186 Val Leu Arg Ile Ala Leu Ser Arg Lys Asp Ile Asn Val Val Ala Val
187           20           25           30
189 Asn Asp Pro Phe Ile Ala Ala Glu Tyr Ala Ala Tyr Met Phe Lys Tyr
190           35           40           45
192 Asp Ser Thr His Gly Arg Tyr Gln Gly Glu Val Thr Phe Glu Gly Lys
193           50           55           60
195 Tyr Leu Val Ile Asp Gly Gln Lys Ile Glu Val Phe Gln Glu Arg Asp
196   65           70           75           80
198 Pro Ala Asp Ile Pro Trp Gly Lys Glu Gly Val Asp Phe Val Ile Asp
199           85           90           95
201 Ser Thr Gly Val Phe Thr Thr Thr Ala Gly Ala Gln Lys His Ile Asp
202           100          105          110
204 Ala Gly Ala Lys Lys Val Ile Ile Thr Ala Pro Ser Ala Asp Ala Pro
205           115          120          125
207 Met Phe Val Met Gly Val Asn His Lys Glu Tyr Thr Lys Asp Leu Ser
208           130          135          140
210 Ile Val Ser Asn Ala Ser Cys Thr Thr Asn Cys Leu Ala Pro Leu Ala
211 145           150          155          160
213 Lys Val Val Asn Asp Val Phe Gly Ile Glu Ser Gly Leu Met Thr Thr
214           165          170          175
216 Val His Ser Ile Thr Ala Thr Gln Lys Thr Val Asp Gly Pro Ser His
217           180          185          190
219 Lys Asp Trp Arg Gly Gly Arg Thr Ala Ser Gly Asn Ile Ile Pro Ser
220           195          200          205
222 Ser Thr Gly Ala Ala Lys Ala Val Gly Lys Val Leu Pro Ala Leu Ala
223           210          215          220
225 Gly Lys Leu Thr Gly Met Ser Leu Arg Val Pro Thr Thr Asp Val Ser
226 225           230          235          240
228 Val Val Asp Leu Thr Val Asn Leu Lys Thr Pro Thr Thr Tyr Ala Glu
229           245          250          255
231 Ile Ser Ala Ala Ile Lys Lys Ala Ser Glu Gly Glu Leu Ala Gly Ile
232           260          265          270
234 Leu Gly Tyr Thr Glu Asp Ala Val Val Ser Thr Asp Phe Leu Thr Asp
235           275          280          285
237 Asn Arg Ser Ser Ile Phe Asp Ala Ser Ala Gly Ile Leu Leu Thr Pro
238           290          295          300
240 Thr Phe Val Lys Leu Ile Ser Trp Tyr Asp Asn Glu Tyr Gly Tyr Ser
241 305           310          315          320
243 Thr Arg Val Val Asp Leu Leu Glu His Val Ala Lys Val Ser Ser Ala
244           325          330          335
247 <210> SEQ ID NO: 7
248 <211> LENGTH: 1491

```

RAW SEQUENCE LISTING

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:15

Input Set : A:\81356224.APP

Output Set : N:\CRF4\01052006\J511436A.raw

249 <212> TYPE: DNA

250 <213> ORGANISM: *Ogataea minuta*

252 <400> SEQUENCE: 7

```

253 aagcttttact ggttcaaggg gttaagtagg ggcgcggtct ggtctttgtg gttgttttcta 60
254 cacggaccac agttgacagc atcgactgct catcgaaaac ggtcgcagtg cggcaatctg 120
255 ctctatctaa tcccaggcta ctcgatccct gcacaaccta cagagtgatc cgaccgcact 180
256 gcccagagatt cagcagactc tcgcagcgca gcgtgcgttt taatccctca aatcaaggct 240
257 gtgcagaccc ggaggatgtg aagctgggac ggcgggaggg aagtctggag tggtagagaga 300
258 atgtgggagc tgtgcaaagg ggcaatggtc actcagcgca gagcgatggg ggcgcggggg 360
259 ccaatatctc ggcaacaaga acgcccagag acgacgggac tctgaatgag agcacgttgt 420
260 ctttcagaca gtccaccgag attccaatat tcgcaggact cgcgctcaga aacgcaaccc 480
261 cggcagattc gcgtccagtc aggccatctg cggcgagctg ctgcgctcgc gggctgcgcc 540
262 acaacgcacg gccacatata cgtcacccgc cgcccgtgg caacctgagg tttttccgca 600
263 acgggtgcac tgattgtgct gttaacgagg caactggaga tgtcagaggc caagtggagc 660
264 catatcacag cggactgcgc atctctggcc tgccggacgc ggtagcgtcc cgtctttttg 720
265 cggacagctt cttaaaacct ggctgaaact aagcgagacc tgcgacctgg aacgcccgca 780
266 caccggtaca cctccggagt tgtatcctca gaagcggagt aacctgcagg cctacgcaag 840
267 aaaagagccc gggacccatc gaccggaaaa gaggggtgga gctagtgggg tagccttgga 900
268 gcagacctgg ggcagacctg ggtagtagc agggccgaaa agggtcagag gaatcagggt 960
269 ggcacggcag tctataccgt agaagctctt ctgcagacga gcgagcagaa actgcacaga 1020
270 ggtccgttcg ccagtctcgt accaccaccg catgacccaa tcagcattga tgctcccaca 1080
271 tgggtagtgc gcgcgaacgc ctggcaccca aacacaccac ttacgcttcc cgcaccgcgg 1140
272 tggttaacac tggcccggag tagtcatata cggagatttt ggcatgattc taattccggg 1200
273 tcgggacacg acctaagtgg cgtgcaaagc tcgggggcta aatgtttccc ggcgctcgcg 1260
274 gcgactcttg tgcgcgcccg cggcggttcg cgggagacgg gggaaagaga ggggtgaccg 1320
275 cagcgagcga tgggtgtgcc gatctcaggc cgagtcaaga caatatataa agagaggatt 1380
276 gtccactttt ctccaatagt atttgaccgc ggttgctctc tgttgatttt ttctagatca 1440
277 tacaattatt gtttgaattc actcaattaa catacacaaa tacaatacaa a 1491

```

280 <210> SEQ ID NO: 8

281 <211> LENGTH: 524

282 <212> TYPE: DNA

283 <213> ORGANISM: *Ogataea minuta*

285 <400> SEQUENCE: 8

```

286 gtggatagat gaccaatggc ctctttaagt aaacatttcg ttttgaatat atttcaagtt 60
287 gaataatgaa agccttggtg tagacttact ccgaagctcc ggggcttcgg ctccctgaat 120
288 ttatttttta catctctgca ccggaaaact ggctatttga aaaatttcga cgttttgctt 180
289 gaaactcgag ttgaggagca ttgcaaatt cgatcgtttt ctaacggacg ccagtcgagt 240
290 tattgttatg tcacgtgaca tcaattgtcc tctattcctt tttggccgat ctcgtttggtg 300
291 ctgacggcct ccgaacagtt acttctaccg gcagggattg gggatgatcg ggatcgatgt 360
292 cctcaactcc agaggctgat ccgatgcggt gggacttcat gcgtccaaat ctgttggtg 420
293 atgtgctctt ctgctttttt ggtgaccaa cgagatgaca attgactgca ttgaaaagg 480
294 tattagcttt tttggtcttc tcctgtgtcg attcgagcgg tacc 524

```

297 <210> SEQ ID NO: 9

298 <211> LENGTH: 113

299 <212> TYPE: DNA

300 <213> ORGANISM: Artificial Sequence

302 <220> FEATURE:

303 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

304 primer for production of an expression cassette

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/511,436A

DATE: 01/05/2006
TIME: 11:08:16

Input Set : A:\81356224.APP
Output Set: N:\CRF4\01052006\J511436A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; N Pos. 3,24,27
Seq#:4; N Pos. 3,6,21,24,27
Seq#:13; N Pos. 3,6,18,21,27
Seq#:14; N Pos. 6,9,15,21,24,27,30
Seq#:25; N Pos. 6,9,12,18,24,27,36
Seq#:26; N Pos. 3,12,15,33
Seq#:38; Xaa Pos. 3,4
Seq#:40; N Pos. 3,21
Seq#:41; N Pos. 19,22,25
Seq#:49; N Pos. 3,12,18
Seq#:50; N Pos. 6,18
Seq#:51; N Pos. 35
Seq#:53; Xaa Pos. 2
Seq#:54; Xaa Pos. 6,9,10
Seq#:55; N Pos. 6,12,18,21,30
Seq#:56; N Pos. 3,6,12,15,18,24,27,30
Seq#:59; Xaa Pos. 1,5
Seq#:61; N Pos. 15,21
Seq#:62; N Pos. 12,27
Seq#:67; N Pos. 3,6,12,15,21,27
Seq#:68; N Pos. 3,12,18,24
Seq#:75; N Pos. 3,6,9,12,15
Seq#:76; N Pos. 6,12,18
Seq#:97; N Pos. 2,5,11
Seq#:98; N Pos. 3,6,12,15,18
Seq#:105; N Pos. 3,6,9,12,15
Seq#:106; N Pos. 1,7,16,19
Seq#:113; N Pos. 6,12,15,21
Seq#:114; N Pos. 9,21

VERIFICATION SUMMARY

DATE: 01/05/2006

PATENT APPLICATION: US/10/511,436A

TIME: 11:08:16

Input Set : A:\81356224.APP

Output Set : N:\CRF4\01052006\J511436A.raw

L:73 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3 after pos.:0
L:112 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0
L:384 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:0
L:433 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14 after pos.:0
L:704 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25 after pos.:0
L:738 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26 after pos.:0
L:998 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:38 after pos.:0
L:1033 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:40 after pos.:0
L:1062 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41 after pos.:0
L:1295 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:49 after pos.:0
L:1319 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:50 after pos.:0
L:1333 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:0
L:1464 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:53 after pos.:0
L:1489 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:54 after pos.:0
L:1529 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:55 after pos.:0
L:1583 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:56 after pos.:0
L:1756 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:59 after pos.:0
L:1791 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:61 after pos.:0
L:1815 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:62 after pos.:0
L:2005 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:67 after pos.:0
L:2039 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:68 after pos.:0
L:2246 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:75 after pos.:0
L:2275 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:76 after pos.:0
L:2974 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:97 after pos.:0
L:3013 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:98 after pos.:0
L:3222 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:105 after pos.:0
L:3256 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:106 after pos.:0
L:3513 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:113 after pos.:0
L:3536 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:114 after pos.:0